# PITFALLS AND POTENTIAL OF SECONDARY DATA ANALYSIS OF THE COUNCIL OF MINISTERS OF EDUCATION, CANADA, NATIONAL ASSESSMENT

*W. Todd Rogers, John O. Anderson, Don A. Klinger, & Teresa Dawber*

Large-scale scale testing agencies, which are generally located within provincial ministries of education and the Council of Ministers of Education, Canada, have called for secondary data analysis of test and survey data that link contextual and educational factors to student performance. First, the potential of secondary analyses is outlined, followed by discussion of seven pitfalls that need to be addressed to realize this potential in Canada. Greater collaboration between testing agencies and teams of qualified secondary analysis researchers needs to take place to ensure that policy issues relevant to education are addressed in meaningful ways.

Key words: large-scale testing, policy analysis, missing data

Les organismes responsables des épreuves communes, qui relèvent généralement des ministères provinciaux de l'Éducation et du Conseil des ministres de l'Éducation (Canada), ont demandé une analyse de deuxième ordre des données d'épreuves et d'enquêtes reliant des facteurs contextuels et pédagogiques au rendement des élèves. Le potentiel des analyses de deuxième ordre est d'abord présenté ; suit une discussion des sept écueils à éviter pour en tirer profit au Canada. Une meilleure collaboration entre les organismes responsables des tests et les équipes de chercheurs spécialisés dans les analyses de deuxième ordre doit être mise en place afin d'assurer que les questions de politiques en matière d'éducation sont abordées de manière pertinente.

Mots clés : épreuves communs, analyse de politiques, données manquantes

Large-scale testing agencies, which in Canada are generally located within provincial ministries of education and the Council of Ministers of Education, have called for secondary data analysis of test and survey data that link contextual and educational factors to student performance. Testing agencies complete the first or primary analysis of the data collected. These analyses generally consist of summary statistics aggregated at student, school, district, and/or provincial levels, and are completed either by personnel within the agencies or by external contractors. However, agency officials believe that the data can be analyzed in different ways and that deeper, more comprehensive analyses that relate variables at the student, class, and school levels to student achievement measures can be conducted. Consequently, they make data available to independent researchers to perform secondary data analyses designed to better answer the original questions or to answer new questions. Such secondary analyses increase the cost effectiveness of testing programs because data then serve a broader spectrum of research questions and researchers. The primary research questions are addressed and also a series of secondary, more complex and deeper questions. Answers to these questions are potentially of greater interest to administrators and officers of testing agencies and the clients they serve. Typically, secondary data analysts, who are located at universities, are independently funded, thereby adding an aura of autonomy and credibility to the secondary results and findings.

It is possible to see secondary data analyses of provincial testing data (e.g., British Columbia, Alberta, and Ontario), national testing data (the literacy, mathematics, writing, and science assessments conducted by the Council of Ministers of Education), and international testing data (e.g., Third International Mathematics and Science Studies [TIMMS] conducted by the International Education Association and the Programme for International Student Assessment [PISA] conducted by the Organization of Economic and Cooperative Development). Statistics Canada encourages secondary data analyses through the provision of data sets in Research Data Centres located at universities across the country, and offers aid and awards to graduate students who perform secondary analyses using Statistics Canada data sets (e.g., National Longitudinal Study of Children and Youth [NLSCY]). Financial support

for such secondary data analysis studies can also come from the Social Sciences and Humanities Research Council in the form of support through the investigator-framed research program (Social Sciences and Humanities Research Council of Canada, 2005)

Experiences with secondary data analyses in Canada have been mixed. Secondary data analysts have found problems with the data they receive, including but not limited to (a) sampling designs that restrict detailed analysis, (b) the failure to collect data on relevant variables needed for statistical modeling, (c) lack of linkages between contextual and achievement data within and between the levels of analysis, (d) missing data due to non-response, (e) inaccurate student responses to both the survey and achievement items, and (f) incomplete or incorrect documentation.[1] Consequently questions are raised about the integrity of the data and the subsequent validity and usefulness of the results and findings.

## OUR POSITION

The authors of this article have completed a series of large-scale secondary analyses of the 2001 Mathematics and 2003 Writing assessments, which are part of the School Assessment and Indicators Project (SAIP) of the Council of Ministers of Education, Canada (CMEC). The primary intents of these secondary analyses were to develop and validate empirical models that relate indicators of educational resources and conditions, home and school traits, to student achievement on significant learning outcomes within a policy-relevant framework. These models can enhance understanding of school performance, particularly if focused on variables that have implications for educational policy and classroom practice. A collateral outcome is the identification of the data collection strategies for improving the quality and utility of future large-scale assessments.

Our experience has suggested that our initial reservations and cautions about secondary data analyses were well founded. However, we are convinced that the basic concept of secondary data analyses of SAIP, and other large-scale data sets, has merit and can contribute to education in Canada. Further, we believe seven issues are central to the success of secondary data analyses. The first three are concerned with

the nature of secondary data analysis outcomes, the need for clearly established links between the primary agency responsible for the assessment and secondary data analysis researchers, and the need for multidisciplinary secondary data analysis research teams with adequate infrastructure support. The remaining four issues deal with technical questions that often arise about the nature and quality of the data for secondary data analyses.

1. *What is the nature of the results from the secondary data analyses of large-scale data sets in education?*

It is important to recognize the nature of secondary data analyses and what they can and cannot do. Studies, such as those available from CMEC, can be best described as explanatory, nonexperimental research (Johnson, 2001). Kerlinger (1986), for example, provided the following definition of nonexperimental research:

> *Nonexperimental research is systematic empirical inquiry in which the scientist does not have direct control of independent variables because their manifestations have already occurred or because they are inherently not manipulable. Inferences about relations among variables are made, without direct intervention, from concomitant variation of independent and dependent variables*. (p. 348; italics in original)

Thus, secondary analyses of data are nonexperimental in that the data have already been collected from naturally occurring groups of students in classes and schools. Consequently, it is not possible to manipulate the conditions and context, even if these variables could be manipulated in practice. Therefore, there is no sense of *cause-and-effect* as found in experimental research (Cook & Campbell, 1979). Instead, secondary data analyses, which identify associational relationships, are explanatory because the analyses can identify only *explanatory factors* that help explain variation in the dependent variable, which, in many cases, is student achievement. As Pascarella and Terenzini (2005, p. 636) pointed out, the best approach for this purpose is causal modeling (e.g., structural equation modeling [Byrne, 1994; Jöreskog & Sörbom, 1993]) and hierarchical linear modeling (Raudenbush & Byrk, 2002). However, the interpretation of the results yielded by these models must be made

carefully to avoid claiming a cause-and-effect relationship between the independent variables and student achievement. Follow-up experimental and quasi-experimental studies need to be undertaken to confirm such tentative cause-and-effect claims. Failing this, a series of carefully considered, related, non-experimental studies including other plausible variables need to be conducted to develop stronger claims of causality (Abelson, 1995; Asher, 1976). Thus, the potential of secondary data analyses lies in the identification of explanatory factors that, with further experimental or quasi-experimental research, cause variation in student achievement.

2. *Are there clearly established links between large-scale testing agencies and researchers who can do secondary data analyses?*

The potential of secondary data analyses will be greatest when close and clear links exist between large-scale testing agencies and secondary analysis teams (Burkhardt & Schoenfeld, 2003; Pascarella & Terenzini, 2005). These links should be established before the design of the testing program and they should provide for a direct route between policy makers, testing specialists responsible for designing and implementing an assessment program, and secondary analysis teams. Through these links, policy issues that could be addressed in the secondary data analyses could be identified prior to data collection. At that time, variables relevant to policy could be identified. With this procedure, other problems, identified below, could be mitigated.

3. *Who should be the secondary researchers?*

As advocated above, secondary data analysis research involves both policy/conceptual and data analysis issues, and should be directed toward improving student learning. Thus, to be most relevant, secondary analysis research teams should be multidisciplinary teams that collectively possess knowledge and skill in each of the following components:

a. policy formulation and implementation,
b. curriculum,

c. instructional procedures and practices,
d. item development and test and questionnaire construction,
e. administration of tests and survey questionnaires,
f. establishing performance standards and setting cut-scores,
g. classical test and item response models,
h. sample designs, sampling, and large-scale statistical analysis procedures that will likely need to incorporate sampling weights and knowledge and skill with the computer programs needed to conduct these analyses; and
i. writing reports and articles for different audiences.

Figure 1 represents a modification of the deliberative research translation model proposed by Willinsky (2001). This model shows the place secondary data analysis teams should occupy in a research ↔ practice system. Inclusion of secondary analysis teams in this model should enhance student learning by focusing on the needs of policy makers and professional development officers, and on improving classroom practice.

*Barriers to collaboration*. Burkhardt and Schoenfeld (2003) identified three barriers that, if not addressed, inhibit the creation of an infrastructure that will lead to realistic collaboration in educational research and development. The three barriers are:

a. significant loss of autonomy, organizational complexities, and the concomitant expenditure of time and energy to establish and maintain collaboration,
b. possible loss of individual status in authorship, and
c. absence of consistent funding to support large R & D teams. (p. 11)

For the first two barriers, Burkhardt and Schoenfeld (2003) suggest that team members adopt the value systems of the natural sciences. These value systems allow team members to receive appropriate credit (e.g., joint recipient of grants/contracts and joint authorship). Without adequate and sustained funding of secondary analysis teams through institutional support and availability of external funds, the potential impact of these analyses will not be realized.

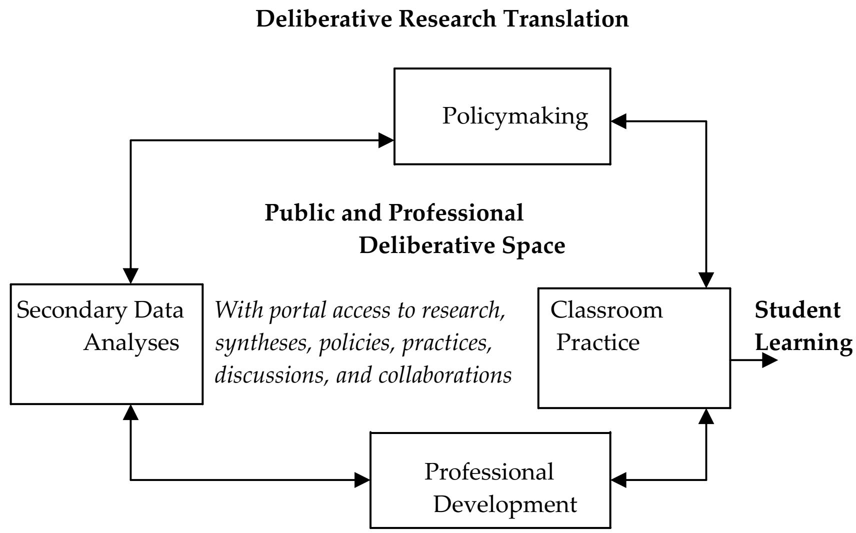


*Figure 1.* *Deliberative research translation for secondary data analysis.*

4. *How representative are the samples of students, teachers, and schools?*

The sample design used by CMEC for selecting students is a three-stage, stratified, unequal probability design. CMEC intends the sample to be a representative sample of students, teachers, and schools for each participating province and for Canada. However, the actual sample is likely not representative because of missing tests and questionnaires and non-response at the item level.

*Missing tests and questionnaires*. The number of students in the database for 13-year-olds for the 2003 CMEC Writing Assessment is 12,708. However, not all students have school and/or teacher data. Deletion of the teachers and schools with missing identification codes, which were to be entered by the students, leaves a sample of 11,570 students.[2] Moreover, an additional 446 students do not have writing scores. Thus, the total number of 13-year-old students with teacher and school information and writing scores is 11,124 (87.5% of the initial sample). Of the 10,972 students assessed in the population of 16-year-olds, 8,663 (80.0%) have teacher and school information and writing

scores. What is not clear is if the students who remain adequately represent the two populations.

*Non-response at the item level*. The data file contained a set of derived scores for school, teacher, and student questionnaires. These scores were determined for the combined 13-year-old and 16-year-old sample using a principal components extraction followed by a varimax rotation of the components with eigenvalues greater than or equal to one (Kaiser, 1970). The number of derived student variables (weighted factor scores) was eight. Only 970 13-year-old students had complete data on these derived variables. Further, only 388 out of the 1,289 teachers of these students had complete data on the 30 derived teacher variables. Clearly, the amount of missing data at the item level is a concern and raises a more serious question about the representation of the student, teacher, and school populations.

*5. What is the integrity of the results?*

Data must have integrity so that the analysis yields results that can be validly interpreted. Incorrectly completing an item and item non-response seriously compromise the integrity of the results. In the Writing Assessment, for example, students recorded their gender twice, once on the cover page and again in response to item 5 in the student questionnaire. Shown in Table 1 is a summary of these two responses for the 13- and 16-year old samples for the Writing Assessment. As can be seen, 24 per cent of students answered the two questions differently (2,951 13-year-old students and 2,533 16-year-old students). Further, there were 154 missing gender responses on the cover page and 976 missing responses for item 5, which resulted in 612 13-year-old students and 492 16-year-old students with missing data for one or both of the gender items. It is not clear which of the two responses, if either, is to be trusted.

*Table 1: Gender Distribution : Cover Page vs. Question 5*

| | Gender: Question 5 | | |
|---|---|---|---|
| Gender: Cover Sheet | Male | Female | Total |
| Male | 4,456 (3,960) | **1,388 (1,118)** | 5,166 (5,078) |
| Female | **1,563(1,415)** | 4,689 (3,987) | 5,589 (5,402) |
| Total | 6,019 (5,375) | 6,077 (5,105) | 12,096 (10,480) |

Note: Numbers for 16-year-old sample in parentheses.

The second example involves the items used to compute the derived score for socio-economic status. The SES index is a linear composite of mother's and/or father's education (question 51) and occupation (question 52) that students reported on the student questionnaire. The number of missing cases for the four items ranged from 2,472 to 3,383 at age 13 and from 2,365 to 2,497 at age 16. Consequently, the SES index was available for only 4,391 (34.6%) 13-year-old students and 6,110 (55.6%) 16-year-old students. There are several possible explanations for the non-response rates for these items. First, some students likely did not know their parents' educational level and/or occupation. Second, some students may have omitted these items because they found it difficult to classify their parent(s)' occupation in the 10 occupational categories.

These examples raise questions about the integrity of the questionnaire data. List wise deletion of cases results in a massive reduction in the number of cases that, in turn, raises questions about what populations the samples represent.

*6. Are the correlations between performance and the derived background variables unexpectedly low?*

Student performance on the Writing assessment is scored using a six-point scoring rubric. The zero-order correlations between the writing scores and the derived variable scores for 13-year-old students ranged from –0.165 to 0.236, with the majority less than $|0.10|$. This can be seen in Table 2: 85.9 per cent of the 13-year-olds and 80.0% of the 16-year olds received scores in three adjacent score categories. This restricted scaling constrains variability, which in turn influences the correlations.

Table 2

*Distribution of Students by Performance Level*

| | 13-year-old | Sample | 16-year-old | Sample |
|---|---|---|---|---|
| Performance Level | *N* | % | *N* | % |
| Below 1 | 653 | 5.1 | 590 | 5.4 |
| Level 1 | 1,991 | 15.7 | 895 | 8.2 |
| Level 2 | 5,199 | 40.9 | 3,012 | 27.5 |
| Level 3 | 3,718 | 29.3 | 4,169 | 38.0 |
| Level 4 | 607 | 4.8 | 1,589 | 14.5 |
| Level 5 | 47 | 0.4 | 261 | 2.4 |
| Missing | 493 | 3.9 | 456 | 4.2 |
| Total | 12,708 | 100.0 | 10,972 | 100.0 |
| Mean | | 2.15 | | 2.58 |
| Standard Dev. | | 0.94 | | 1.10 |

7. *Can more advanced statistical procedures be used with the present CMEC sample design?*

Organizations like CMEC would like secondary analysis teams to look at deeper, more complex policy relevant questions.[3] The structure of national assessment data sets, such as the CMEC data sets, is typically students nested within classes nested within schools nested within provinces. The Hierarchical Linear Model (HLM) (Raudenbush & Byrk, 2002; Luke, 2004) is a statistical procedure that can separate school effects from class effects from student effects. If these level effects are not separated, then student findings could be incorrect. Given the complex interrelationships among student, class, and school variables, HLM is superior to ordinary least squares regression procedures.

HLM requires that sample sizes at each level be relatively large to allow estimation of effects with small sampling error. Although some agreement exists that the number of higher level units like schools should be at least 30 (Snijders & Bosker, 1999), little agreement exists on the required number of individuals or students within each unit. Willms (1992) suggested that estimates for schools with fewer than 20 students are not very stable. For the 16-year-old writing sample, there are only 13 school/class combinations with 20 or more students. Consequently, it is not possible to generalize the results yielded by the HLM analyses to the full population represented by the initial samples of schools and students.

Obviously a better sampling procedure is required. For example, the potential for secondary data analyses of CMEC data would be enhanced if the lowest sampling unit were changed from student to class. That is, classes within schools should be selected and all students in the selected classes chosen. Some school co-coordinators indicated that it would be less disruptive if intact classes of students were assessed (Council of Ministers of Education, 2002b, p. 49). As pointed out by Willms (1992), estimates of school and class effects are needed in addition to student effects (p. 49). However, it likely will be necessary to use a 2-level HLM because many schools have only one class, thereby preventing the separation of class and school effects.

## CONCLUSION

We have discussed what we believe to be major pitfalls that work to inhibit the potential of secondary data analyses of CMEC assessment data sets. However, we want to emphasize that with appropriate changes, the potential of the CMEC assessment program to inform, in telling and meaningful ways, policy makers and users of the CMEC results can be realized. To effect these changes, we believe that a closer link needs to occur between CMEC and secondary analysis researchers. These researchers need to work with the policymakers, testing specialists, and analysts of the large-scale testing agency during the design and planning phase. They are in a position to pose significant questions and to identify and help ensure that a) the variables to be considered are relevant to the improvement of classroom practice and student learning; b) assessment instruments, background questionnaires, and scoring of responses to these instruments will yield scores that reflect the concomitant variation expected for the independent and dependent variables; and c) that the samples are representative of the populations to which the results are to be generalized and of sufficient size to permit the use of the complex data analysis procedures that yield stable, unbiased results needed to answer the questions posed. This collaboration should be on-going, beginning with the formulation of questions and the design of these programs, followed by instrument development, sample design, and scoring, and the analysis phase as analytic models are developed, evaluated, and refined. We believe that the following question should shape the outcomes of this collaboration: *How can the use of secondary data analysis results be increased in schools and school districts, thereby leading to improved student learning mediated through the informed practice of school leaders and classroom teachers*?

## NOTES

[1] Anderson, Rogers, Klinger, Ungerlieder, Glickman, and Anderson (2006) found similar problems in their analyses of the 2001 CMEC Mathematics Assessment. Likewise, Pascarella, and Terenzini (2005), in their analyses of how college education affects students, identified similar problems and issues in their examination of the secondary data sets they worked with.

[2] Students who were not currently studying English Language Arts were instructed to leave the teacher ID blank, accounting for some of the missing identification numbers (Council of Ministers of Education, Canada, 2002a).

[3] The Canada Education Statistics Council (CESC) and the Social Sciences and Humanities Research Council (SSHRC) through the Education Research Initiative (ERI) clearly identified education, and learning outcomes in particular, as high priority areas for research. ERI intended to stimulate empirical research and quantitative analysis to link available databases containing student assessment information to educational policy issues in Canada. However, CESC and SSHRC no longer support ERI.

REFERENCES

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. O., Rogers, W. T., Klinger, D. A., Ungerlieder, C., Glickman, V., & Anderson, B. (2006). Student and School Correlates of Mathematics Achievement: Models of School Performance Based on Pan-Canadian Student Assessment. *Canadian Journal of Education*, 29(3), 706-730.

Asher, H. B. (1976). *Causal modeling*. University Paper series on Quantitative Applications in the Social Sciences, 07-003. Beverly Hills & London, UK: Sage Publications.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, *32*(9), 3-14.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage Publications.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally.

Council of Ministers of Education, Canada. (2002a). School achievement indicators program (SAIP), Writing III Assessment, Handbook for Schools. Toronto: Author.

Council of Ministers of Education, Canada. (2002b). School achievement indicators program (SAIP), Writing III Assessment, 2002, Technical Report. Toronto: Author.

Johnson, B. (March, 2001). Toward a new classification of nonexperimental research quantitative research. *Educational Researcher*, *30*(2), pp. 3-13.

Jöreskog, K. G., & Sörbom, D. (1993). *New features in LISREKL 8*. Chicago: Scientific Software.

Kaiser, H. (1970). A second generation Little Jiffy. *Psychometrika*, *35*, 401-415.

Kerlinger, F. N. (1986). *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.

Luke, D. A. (2004). Multilevel modeling. Sage university paper series. Quantitative applications in the social sciences, 07-143. Beverly Hills, CA: Sage Publications.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (2nd ed.). San Francisco: Jossey-Bass.

Raudenbush, S. W., & Byrk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Social Sciences and Humanities Research Council of Canada. (2005). Knowledge council: SSHRCC, 2006-2011. Ottawa: Author.

Snijders, T., & Bosker, J. B. (2003). Modeled variance in two-level models. *Sociological Methods & Research*, *22(3)*, 342-363.

Willinsky, J. (2001). The strategic education research program and the public value of research. *Educational Researcher*, *30*(1), 5-14.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: The Falmer Press.